



Logistiek algemeen

Forecastkwaliteit verbeteren? Automatiseer waar mogelijk en voorspel zelf waar nodig

Het correct voorspellen van klantvraag is belangrijk om betrouwbare planningen te maken en beslissingen te nemen.

62

Nick van de Bunt en Nienke Hofstra

HAN University of Applied Sciences

Het correct voorspellen van klantvraag is belangrijk om betrouwbare plannings te maken en beslissingen te nemen. Geautomatiseerde systemen kunnen voorspellers ondersteunen met statistische voorspellingen. Deze kunnen menselijke beoordelaars vaak aanpassen. Aanpassingen kunnen voorspellingen verbeteren, maar zijn niet altijd waardevol. In dit project hebben we het forecastproces bij een Nederlandse fabriek onderzocht en gekeken waar en hoe deze geautomatiseerd kan worden. Hiervoor is een analyse gemaakt van de kwaliteit van de voorspellingen en gemaakte aanpassingen. Het advies is om het forecastproces te automatiseren waar mogelijk (stabiele vraagpatronen), en het menselijk oordeel in te zetten waar nodig (complexere vraagpatronen).

Introductie

Het correct voorspellen van klantvraag is belangrijk om betrouwbare plannings te maken en activiteiten uit te voeren. Hiermee kunnen juiste beslissingen genomen worden over het doen van investeringen, het aanhouden van voorraden, productieniveaus en inkoopvolumes. Zo kan er beter worden ingespeeld op de klantvraag, kunnen voorraadkosten gereduceerd worden, levertijden verkort worden en resources efficiënter worden ingezet (Boone et al., 2019).

Het maken van goede vraagvoorspellingen is echter niet eenvoudig. Zeker niet wanneer er veel vraagvoorspellingen moeten worden gemaakt. In de praktijk zijn daarom geautomatiseerde systemen beschikbaar die planners, accountmanagers of andere functionarissen kunnen ondersteunen bij het maken van prognoses. Deze systemen bevatten rekenmethodes die aan de hand van historische data een statistische voorspelling maken.

In de praktijk worden deze geautomatiseerde systemen echter niet altijd gebruikt. En als ze gebruikt worden, worden ze over het algemeen ook door mensen beoordeeld en indien nodig geacht, aangepast (Siemsen & Aloysius, 2019). Deze aanpassingen kunnen de kwaliteit van voorspellingen verbeteren omdat mensen over impliciete kennis (tacit knowledge) kunnen beschikken die het systeem niet kent, zoals informatie over acties van concurrenten of marketingcampagnes (Sanders & Manrodt, 2003).

Maar niet alle aanpassingen zijn waardevol (Trapero et al., 2013). Er zijn verschillende ‘behavioral biases’ – systematische afwijkingen van een bepaalde norm, zoals systeemvoorspellingen – die voorspellingen kunnen beïnvloeden. Het gevoel van overprecisie (het overschatten van nauwkeurigheid van je informatie) dat mensen vaak hebben, leidt er bij het maken van prognoses bijvoorbeeld vaak toe dat de werkelijke variantie wordt onderschat (Eroglu & Croxton, 2010). Bovendien is bekend dat mensen de neiging hebben om algoritmes te wantrouwen en daarom af te wijken van systeemvoorstellen (Dietvorst et al., 2015).

Om als organisatie een beleid te formuleren over de inrichting van een forecastproces en over de rol die menselijke beoordelingen daarin zouden moeten spelen, is het dus belangrijk om te begrijpen wanneer mensen systeemvoorspellingen aanpassen en of en wanneer die aanpassingen de kwaliteit van de voorspellingen verbetert. Op dit vlak zijn diverse onderzoeken uitgevoerd. Zo zijn er gedragstendensen geïdentificeerd die de kwaliteit van voorspellingen beïnvloeden (Kremer et al., 2011), is onderzocht wat voor type aanpassingen de kwaliteit van voorspellingen verbeteren (Fildes et al., 2009), en is bekeken welke rol productkarakteristieken spelen in het maken van systeemaanpassingen (Khosrowabadi et al., 2022).

64

In dit project onderzoeken we het forecastproces en de kwaliteit ervan bij The Dutch Factory (geanonimiseerde naam; hierna TDF). TDF ervaart een lage forecastkwaliteit. Op dit moment maakt TDF geen gebruik van geautomatiseerde forecastsystemen maar maken accountmanagers ieder op hun eigen wijze vraagvoorspellingen voor soms meer dan 300 artikelnummers. We onderzoeken hoe TDF geautomatiseerde technologie en expertise van accountmanagers in optimale combinatie kan inzetten om de kwaliteit van de vraagvoorspellingen te verbeteren.

Hoe kunnen mens en systeem samenwerken om de kwaliteit van vraagvoorspellingen te verbeteren?

Hiertoe bestuderen we allereerst de literatuur om kennis op te halen over de inzet van rekenmethodes om vraagvoorspellingen te maken, over wanneer aanpassingen worden gemaakt en het effect van aanpassingen op de kwaliteit van voorspellingen (H2). Daarna leggen we uit hoe we onze analyse hebben uitgevoerd (H3). Vervolgens beschrijven we de resultaten (H4) gevolgd door het advies aan TDF (H5). We eindigen met een conclusie (H6).

Theoretische achtergrond

Om advies te kunnen formuleren aan TDF over de inzet van geautomatiseerde systemen en welke rol mensen daarin het beste kunnen spelen, is het belangrijk om te begrijpen (1) wanneer mensen systeemvoorspellingen aanpassen en (2) wat de impact daarvan is op de kwaliteit van voorspellingen.

Omdat TDF op dit moment geen geautomatiseerd systeem gebruikt om vraagvoorspellingen te maken, brengen we allereerst in kaart welke rekenmethodes er voor handen zijn om vraagvoorspellingen te maken en onder welke condities ze het beste toegepast kunnen worden. We houden deze tegen het licht van de praktijk bij TDF om advies te formuleren over het automatiseren van vraagvoorspellingen.

Om richting te geven aan het onderzoek bekijken we daarna wat er al bekend is over de condities waaronder mensen systeemvoorspellingen aanpassen en over het effect van aanpassingen op de kwaliteit van de voorspellingen. Deze theorieën testen we bij TDF. De uitkomsten zijn input voor het advies over menselijke interventies om de kwaliteit van de vraagvoorspellingen te verbeteren.

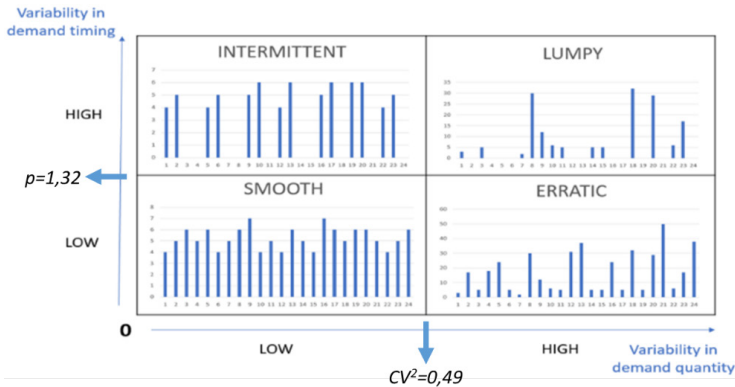
Rekenmethodes voor vraagvoorspellingen

Om toekomstige vraag naar een product in kaart te brengen, kunnen historische verkoopdata worden geanalyseerd. Er zijn diverse rekenmethodes beschikbaar om middels die data voorspelling te genereren. Echter, niet elke methode is altijd geschikt. Het kiezen van de juiste voorspelmethode is belangrijk voor de betrouwbaarheid van de gemaakte voorspellingen.

Welke voorspelmethode geschikt is, hangt af van het vraagpatroon van het artikel. Het vraagpatroon bepaalt hoe 'voorspelbaar' een artikel is en hangt af van twee factoren (Syntetos et al., 2005):

p , het gemiddelde tijdsinterval tussen twee vraagmomenten, berekend als CV^2 , de variatie in de omvang van de vraag, gemeten als

De mate van variabiliteit in het tijdsinterval en de vraag grootte (hoog of laag) bepaalt het vraagpatroon: de afkapwaarden voor variabiliteit in het tijdsinterval en de vraag grootte zijn respectievelijk 1,32 en 0,49. Dit resulteert in vier vraagpatronen: 'smooth', 'intermittent', 'erratic' en 'lumpy'. Ze zijn gevisualiseerd in figuur 1 hieronder.



Figuur 1 Vraagpatronen

Opmerking. Overgenomen uit *Demand classification: Why forecastability matters* (p. 1) door Frepple.com, z.d. (<https://frepple.com/blog/demand-classification/>). Copyright z.d., Frepple.com.

66

Producten met een 'smooth' vraagpatroon kunnen volgens de literatuur het beste voorspeld worden aan de hand van de rekenmethode Moving Average of Exponential Smoothing (Ragnerstam, 2015). De Moving Average methode houdt in dat de gemiddelde klantvraag over een bepaalde periode wordt berekend waarbij de periode steeds opschuift (wordt geactualiseerd). Exponential Smoothing is gebaseerd op hetzelfde principe, alleen krijgt recentere data een zwaardere gewicht. Het gewicht van de voorgaande periodes wordt bepaald door α . De α ligt tussen 0 en 1 en de waarde kan een organisatie zelf bepalen. In de literatuur wordt een waarde tussen 0,1 en 0,3 geadviseerd (Syntetos et al., 2005).

Producten met een 'intermittent' vraagpatroon kunnen het beste voorspeld worden volgens de Crostons methode (Fu et al., 2018). Intermittent vraagpatronen kennen periodes waarin de afzet nul is. In een Moving Average of bij Exponential Smoothing tellen deze periodes mee, wat leidt tot een te lage voorspelling wanneer er wel klantvraag is. Croston's methode werkt met Exponential Smoothing, maar neemt maanden waarin de afzet 0 was niet mee. Ook Exponential Smoothing kan, met een andere α , volgens de theorie een geschikte voorspelmethode zijn (Talele, 2020).

Producten met een 'erratic' vraagpatroon kennen een non-lineair patroon met niet periodiek-specifieke variaties. Dit leidt tot grote onzekerheden. Het voorspellen van deze vraagpatronen is dan ook lastig (Jiang et al., 2017). Dit vraagpatroon komt vooral voor bij halffabricaten (ook TDF produceert halffabricaten). Het non-lineaire patroon maakt methodes zoals Exponential Smoothing en Croston's methode ongeschikt. Complexere machine learning methodes zijn beschikbaar om voorspellingen te genereren maar deze vragen veel expertise en rekenkracht bij toepassing (Erasmus University Rotterdam, 2018).

'Lumpy' vraagpatronen zijn vanwege de gecombineerde variabiliteit in tijdsinterval en vraag grootte nauwelijks te voorspellen met een rekenmethode.

In hoofdstuk 4 bepalen we het vraagpatroon onderzoeken we welke rol het vraagpatroon speelt in de forecastkwaliteit bij TDF. In hoofdstuk 5 onderzoeken we welke rekenmethode het beste gebruikt kan worden voor het voorspellen van welke vraagpatronen bij TDF.

Aanpassingen in systeemvoorspellingen

Er zijn diverse onderzoeken uitgevoerd naar de interactie tussen mensen en geautomatiseerde forecastsystemen. Onderzoeken richten zich enerzijds op het identificeren en begrijpen van de beweegredenen om aanpassingen te maken. Denk hierbij bijvoorbeeld aan onderzoek naar zogenaamde 'behavioral biases', systematische afwijkingen van systeemvoorspellingen. Zo hebben De Baets en Harvey (2018) gezien dat mensen de neiging hebben om tijdens promotieperiodes de vraag te laag in te schatten en tijdens normale periodes te hoog. Kremer et al. (2011) en Eroglu en Croxton (2010) zagen dat mensen de neiging hebben om te kleine aanpassingen in de voorspelling te maken in een omgeving met een onstabiel vraagpatroon en te grote aanpassingen te maken in een omgeving met een stabiel vraagpatroon.

67

Naast het vraagpatroon zijn ook de rol van andere productkarakteristieken in forecasting onderzocht. Hierbij kun je denken aan zaken als de productclassificatie (A, B en C producten), productprijs, 'fast-movers' versus 'slow-movers' en de productcategorie. Khosrowabadi et al. (2022) hebben een grote dataset van een Europese retailer onderzocht om te begrijpen wanneer en hoe mensen de accuraatheid van systeemvoorspellingen kunnen verbeteren. Ze vonden dat planners slechts een fractie van de voorspellingen aanpaste. Planners pasten producten met een groter belang voor het bedrijf (A klasse) vaker aan dan producten met een kleiner belang voor het bedrijf.

Onderzoeken richten zich anderzijds op het in kaart brengen van de impact van menselijke aanpassingen op de kwaliteit van voorspellingen. Kremer et al. (2011) zagen dat de kwaliteit van voorspellingen slechter is in stabiele omgevingen dan in instabiele omgevingen. In hun studie naar vraagvoorspellingen in retail vonden Khosrowabadi et al. (2022) dat slechts de helft van de aanpassingen de vraagvoorspelling verbeterde. Verder hebben ze gezien dat ook de richting en magnitude van de aanpassing een belangrijke rol spelen in de kwaliteit van de gemaakte aanpassing. Grote positieve aanpassingen werden vaker gedaan maar waren relatief vaker fout dan de grote negatieve aanpassingen. Dit is in lijn met bevindingen van Fildes et al. (2009).

In hoofdstuk 4 gaan we na welke rol de productkarakteristieken classificatie en vraagpatroon spelen in de kwaliteit van voorspellingen en bij het maken van systeemaanpassingen. Daarbij gaan we na of belangrijke artikelen inderdaad vaker aangepast worden dan producten met een kleiner belang voor het bedrijf (de impact van productclassificatie), of de kwaliteit van voorspellingen slechter is in stabiele omgevingen dan in instabiele omgevingen (de impact van vraagpatronen) en wat de impact is van grote versus kleine positieve en negatieve aanpassingen op de kwaliteit van voorspellingen.

Methodologie

In dit project onderzoeken we hoe mens en systeem kunnen samenwerken om de kwaliteit van vraagvoorspellingen bij The Dutch Factory te verbeteren. TDF produceert producten die een hoofdrol spelen in de productieprocessen van andere ondernemingen. De belangrijkste producten worden op voorraad geproduceerd. Om te bepalen hoeveel er geproduceerd moet worden, worden vraagvoorspellingen gemaakt voor alle producten. De producten zijn ingedeeld in drie groepen: S (Standard), C (Commitment) en N (Non-standard). De S- en C-producten worden make-to-stock geproduceerd. De belangrijkste producten binnen de standaardproducten betreft de 'B40' producten. Dit betreft 20% van alle producten, maar is samen goed voor ongeveer 80% van de omzet van TDF. In dit project focussen we ons specifiek op deze B40 producten.

68

Het forecast proces verloopt handmatig, zonder gebruik te maken van geautomatiseerde systemen die ondersteuning bieden door het doen van statistische voorspellingen. De 'forecast accuracy' (het aantal of aandeel voorspellingen dat gelijk is aan de vraag) van de B40 producten is gemiddeld over de afgelopen 12 maanden 49%. Om de negatieve gevolgen hiervan, zoals de hoge (of lage) voorraadstanden van verkeerde producten en de daarmee gemoeide kosten, te verminderen is het van belang de forecast kwaliteit (hoe dicht de voorspellingen de vraag benaderen) te verbeteren. Hieronder geven we een nadere beschrijving van hoe het voorspellen van de vraag bij TDF in zijn werk gaat. Daarna beschrijven we onze onderzoeks aanpak om de kwaliteit van de vraagvoorspellingen nader te onderzoeken en te bekijken hoe geautomatiseerde systemen ondersteuning kunnen bieden.

Beschrijving van het demand forecasting proces

Binnen TDF worden vraagvoorspellingen van klantproducten gemaakt om productieplanningen te realiseren. Het maken van vraagvoorspellingen bestaat uit twee stappen: de accountmanagers maken een voorspelling van de vraag naar producten van klanten uit hun portfolio en deze worden door het demand review team gecontroleerd en eventueel gewijzigd. Beide stappen worden meegenomen in dit onderzoek.

Bij het bedrijf werken 14 accountmanagers. Zij beheren elk hun eigen klantportfolio en maken maandelijks demand forecasts voor hun klantproducten. Daarbij wordt 6 maanden vooruit geforecast. Daarom zijn steeds preliminaire forecasts voor de komende 5 maanden zichtbaar. Accountmanagers kunnen deze wijzigen en vullen een forecast in voor de zesde maand. Zo wordt (in theorie) elke maand 6 keer geëvalueerd. Over het algemeen wordt, naast het invullen van maand 6, vooral gekeken naar de forecast voor de komende maand. Deze is input voor de productieplanning voor de komende maand en daarom het belangrijkste voor de accountmanagers. In ons onderzoek richten we ons daarom specifiek op de vraagvoorspellingen voor komende maand (latere maanden blijven buiten beschouwing).

De accountmanagers hebben één week de tijd om een invulbestand te vullen waarin ook historische salesdata met gemiddelden van de afgelopen 12, 9, 6 en 3 maanden per klantartikel zijn opgenomen. Afhankelijk van het klantportfolio kan een invulbestand meer dan 300 regels met klantproducten bevatten. Met een forecast horizon van 6 maanden kan dit een grote taak opleveren. Dit wordt versterkt doordat er geen (niet systematisch) gebruik gemaakt wordt van forecasting methoden. Accountmanagers hebben ieder hun eigen werkwijze om een forecast te bepalen. Eén accountmanager gebruikt een eigen ontworpen rolling forecast, anderen kijken naar data uit het verleden en maken op basis van marktkennis en gevoel een inschatting. De maandelijkse forecasts die door de accountmanagers wordt opgeleverd, wordt FC0 genoemd.

69

Deze forecasts worden door het demand review team, bestaande uit de general manager en sales manager, gecontroleerd. Zij kunnen wijzigingen aanbrengen en leveren vervolgens de definitieve demand forecast op, genaamd FC1. Na een check door het supply review team en Sales en Operations Planning (S&OP) gebruikt de afdeling SCM deze definitieve forecast om een productieplanning te maken.

De onderzoeksaanpak

Gebruikte datasets voor forecastanalyse

We richten ons in dit project specifiek op de vraagvoorspellingen van de B40 producten. Deze zijn in 2021 goed voor bijna 80% van de totale afzet en daarmee het meest belangrijk voor TDF. We onderzoeken de voorspellingen die zijn gemaakt in de periode maart 2021 tot en met februari 2022 (12 maanden). Voor de analyse van accountmanagement en demand review gebruiken we verschillende datasets. De dataset van accountmanagement bestaat uit de voorspellingen per klantvraag. De dataset bevat 3566 voorspellingen van 286 unieke producten van 11 accountmanagers die ieder per maand gemiddeld 27,0 voorspellingen van B40 producten maken. (Er zijn 53 forecasts met een negatieve vraagrealisatie en 23 forecasts met ontbrekende SKU details. Deze 76 regels zijn weggelaten voor verdere analyse.)

Omdat dezelfde producten door verschillende klanten kunnen worden afgenomen, zijn er producten waarvoor meerdere accountmanagers vraagvoorspellingen maken (gemiddeld 2,4 accountmanagers per product), zij het voor verschillende klantvraag. De FC0 voorspellingen van de accountmanagers worden geaggregeerd op productniveau voor beoordeling door demand review. De gebruikte dataset van demand review bestaat uit 1473 forecasts van 277 unieke producten met gemiddeld 122,8 vraagvoorspellingen per maand. (Er zijn 26 forecasts met een negatieve vraagrealisatie en twee forecasts met ontbrekende SKU details. Deze 28 regels zijn weggelaten voor verdere analyse.)

Om de rol van productklasse en -vraagpatroon in de forecastkwaliteit en -aanpassingen te onderzoeken, nemen we de door TDF gehanteerde ABC classificering van producten over. Het vraagpatroon van producten was vooraf onbekend. We hebben salesdata van januari 2021 tot en met mei 2022 (17 maanden) gebruikt om de vraagpatronen van de producten in onze datasets te bepalen volgens de classificering van Syntetos et al. (2005) (zie sectie 2.1). Omwille van mogelijke impact van de Coronapandemie op salesdata is een vergelijking gemaakt met salesdata uit 2018 en 2019. De voorgenoemde dataset van 2021-2022 bleek representatief voor verkoopdata en is daarom gebruikt voor het identificeren van de vraagpatronen.

70

Aanvullende informatie

Om het huidige forecastproces in kaart te brengen zijn diverse interviews gedaan met de verschillende afdelingen die bij het forecastproces betrokken zijn. Er is gesproken met de manager ICT, de manager Finance, de manager Sales, de manager Supply Chain, de General Manager en met 6 accountmanagers. Daarnaast zijn de verschillende stakeholders geobserveerd tijdens het uitvoeren van het forecastproces om een nog beter beeld te krijgen van werkwijzen.

Resultaten

In dit project onderzoeken we hoe mens en systeem kunnen samenwerken om de kwaliteit van vraagvoorspellingen bij The Dutch Factory te verbeteren. We onderzoeken allereerst in meer detail de huidige kwaliteit van vraagvoorspellingen van zowel accountmanagement (FC0) en demand review (FC1). We onderzoeken hierbij de rol van productkarakteristieken (vraagpatroon en classificatie). Omdat demand review initiële voorspellingen van accountmanagement kan aanpassen, onderzoeken we ook de kwaliteit van gemaakte aanpassingen. En we onderzoeken karakteristieken van aanpassingen. De uitkomsten zijn input voor het advies over menselijke interventies om de kwaliteit van de vraagvoorspellingen te verbeteren.

Om de kwaliteit van vraagvoorspellingen inzichtelijk te maken, onderscheiden we twee zaken. Allereerst kijken we naar de accuraatheid van voorspellingen, waarmee we bedoelen of de voorspelling (FC) gelijk is aan de vraag (D): $FC=D$. Accuratesse alleen zegt niet zo veel, immers foutieve aanpassingen kunnen verschillen in magnitude. Voor de voorspellingen die niet accuraat zijn, bekijken we daarom ook de forecastafwijking, hoe dicht de voorspelling de vraag benadert. Dit berekenen middels de verhouding van de absolute afwijking van de voorspelling (FC_E) tot de vraag: FC_E/D .

Kwaliteit vraagvoorspellingen accountmanagement (FC0)

Forecastkwaliteit op afdelingsniveau

Allereerst hebben we gekeken naar de accuraatheid van de voorspellingen van de accountmanagement afdeling als geheel. Hier wordt bedoeld de accuraatheid van de voorspellingen van de 11 accountmanagers samen. (De afdeling heeft meer accountmanagers, maar zij hebben dermate weinig B40 producten in hun portfolio, of hun klanten bestellen deze dermate sporadisch, dat het opnemen van forecastdata van deze accountmanagers een vertekend beeld zou geven van hun prestaties.) Van de 3566 voorspellingen is slechts 7,0% correct.

Om een beeld te krijgen van de rol van productkarakteristieken is ook een onderscheid gemaakt tussen aandeel correcte voorspellingen per productklasse (A, B of C) en per vraagpatroon (smooth, intermittent, lumpy en erratic). De linkerzijde van onderstaande tabel laat het percentage juiste voorspellingen per productcategorie zien. We willen weten van welke productcategorie de vraag relatief vaker correct wordt voorspeld, daarom corrigeren we hier voor het absolute aantal producten per categorie (en berekenen we geen percentage van het totaal aantal juiste voorspellingen). Het zijn B klasse producten en producten met een erratic vraagpatroon waarvan de vraag het vaakst correct wordt voorspeld. Dat de vraag naar producten met een smooth vraagpatroon niet het vaakst correct worden voorspeld is interessant omdat dit wel het meest stabiele vraagpatroon is. Erratic vraagpatronen kennen juist veel variëteit in de omvang van de vraag.

Tabel 1 Percentage juiste/overschatte/onderschatte FC0 voorspellingen per productcategorie (de dataset kent geen C producten met een smooth vraagpatroon)

	% vraagvoorspellingen correct (FC0=D)				% vraagvoorspellingen overschat (FC0>D)				% vraagvoorspellingen onderschat (FC0<D)			
	A	B	C	Totaal	A	B	C	Totaal	A	B	C	Totaal
Erratic	8,3%	12,2%	3,7%	8,7%	47,5%	45,0%	57,9%	48,2%	44,2%	42,8%	38,3%	43,1%
Intermittent	7,5%	4,6%	6,7%	6,2%	49,3%	52,5%	57,0%	54,2%	43,3%	42,9%	36,2%	39,6%
Lumpy	0,0%	8,6%	6,0%	6,5%	64,9%	48,3%	52,8%	52,0%	35,1%	43,1%	41,2%	41,5%
Smooth	6,3%	8,6%		6,6%	51,7%	47,0%		51,0%	42,0%	44,4%		42,3%
Totaal	7,0%	8,0%	6,1%	7,0%	50,3%	48,8%	55,6%	51,4%	42,7%	43,2%	38,2%	41,6%

Tabel 1 maakt ook een onderscheid tussen voorspellingen die de vraag overschatten en onderschatten (zie het midden en rechts van de tabel). De vraag naar producten wordt vaker overschat dan onderschat (51,4% versus 41,6%, respectievelijk). Dit geldt voor de verschillende productklassen en vraagpatronen. Verder is te zien dat als de vraag naar producten wordt overschat, dit relatief het vaakst gebeurt voor de minst belangrijke producten (C klasse). Als de vraag wordt onderschat zijn het relatief vaker A en B klasse producten. Als we naar het vraagpatroon kijken, zien we dat smooth vraagpatronen, hoewel die het meest stabiel zijn, niet het minst vaak wordt overschat, noch onderschat. En hoewel lumpy vraagpatronen de meeste variatie kennen, worden die ook niet het vaakst overschat of onderschat. Deze bevindingen zijn in tegenstelling tot eerder onderzoek waarin is gevonden dat het juist stabiele (smooth) vraagpatronen relatief vaker worden overschat.

Van de foute voorspellingen hebben we ook de forecastafwijking geanalyseerd. Onderstaande tabel 2 presenteert per productcategorie de gemiddelde procentuele forecastafwijking van de vraag bij een te hoge voorspelling (links) en een te lage voorspelling (rechts). Dat de te hoge voorspellingen relatief gezien een stuk verder boven de vraag liggen dan dat te lage voorspellingen onder de vraag liggen, wordt mede verklaard door het feit dat een afwijking naar beneden niet meer dan 100% van de vraag kan afwijken.

Tabel 2 Gemiddelde kwaliteit van FC0 per productcategorie

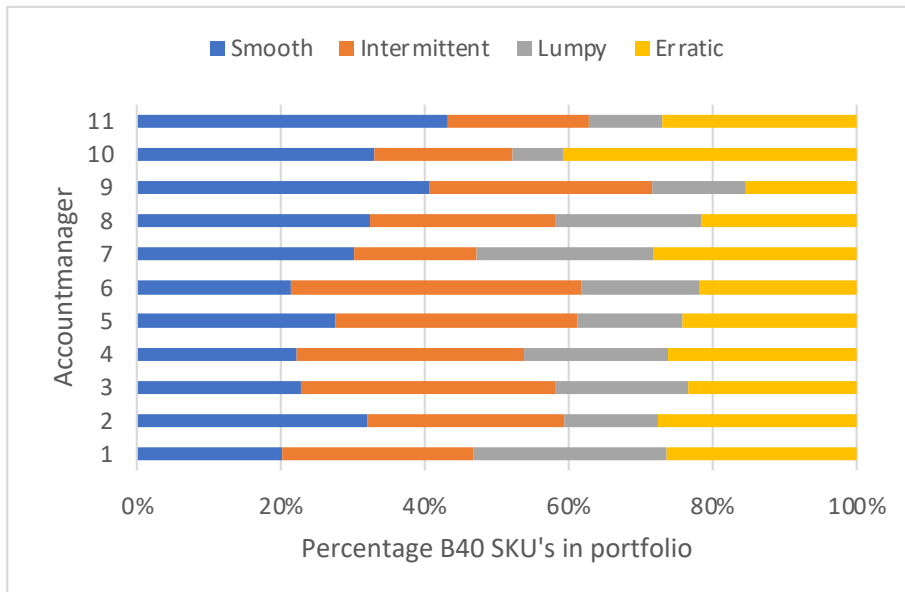
	Gemiddelde procentuele afwijking van overschatte vraagvoorspellingen (FC0>D)				Gemiddelde procentuele afwijking van onderschatte vraagvoorspellingen (FC0<D)			
	A	B	C	Totaal	A	B	C	Totaal
Erratic	341%	598%	446%	416%	-65%	-74%	-76%	-68%
Intermittent	484%	339%	187%	282%	-65%	-59%	-81%	-70%
Lumpy	720%	227%	227%	264%	-63%	-84%	-88%	-85%
Smooth	294%	180%		278%	-55%	-58%		-56%
Totaal	345%	346%	229%		-60%	-68%	-83%	

Opvallend is dat de gemiddelde positieve afwijking van de vraag (FC0>D) het grootst is voor A en B klasse producten terwijl de gemiddelde negatieve afwijking (FC0<D) het grootst is voor C klasse producten. Met andere woorden, belangrijkere producten worden, relatief tot de vraag gezien, het meest overschat en minder belangrijke producten het meest onderschat. Verder is te zien dat de gemiddelde positieve afwijking het grootst is voor producten met een erratic vraagpatroon maar het minst voor producten met een lumpy vraagpatroon. Dit is opvallend omdat beide vraagpatronen grote variabiliteit in de omvang van de vraag kennen. Voorspellingen die te laag zijn, wijken ten opzichte van de vraag dan wel weer het meest af voor producten met een lumpy vraagpatroon.

Als we tabel 1 en 2 met elkaar vergelijken, valt ook iets interessants op. Het vaakst of het minst vaak overschatten of onderschatten van de vraag, zegt niets over de procentuele afwijking van de vraag. Als we bijvoorbeeld kijken naar het aantal te hoge voorspellingen, zien we dat dit het vaakst C klasse producten zijn (tabel 1) terwijl de gemiddelde afwijking van die voorspellingen ten opzichte van de vraag het laagst is (tabel 2). Tegelijkertijd wordt de vraag naar C klasse producten relatief het minst vaak onderschat (tabel 1), maar is de gemiddelde afwijking van de vraag in het geval van onderschatting wel het grootst (tabel 2). Als we kijken naar het vraagpatroon zien we dat intermittent vraagpatronen relatief het vaakst worden overschat (tabel 1), maar dat de omvang van de overschatting een stuk lager is (tabel 2): erratic vraagpatronen worden, in verhouding tot de vraag gezien, veel sterker overschat. Bij onderschattingen zijn het de lumpy vraagpatronen die het sterkst worden onderschat.

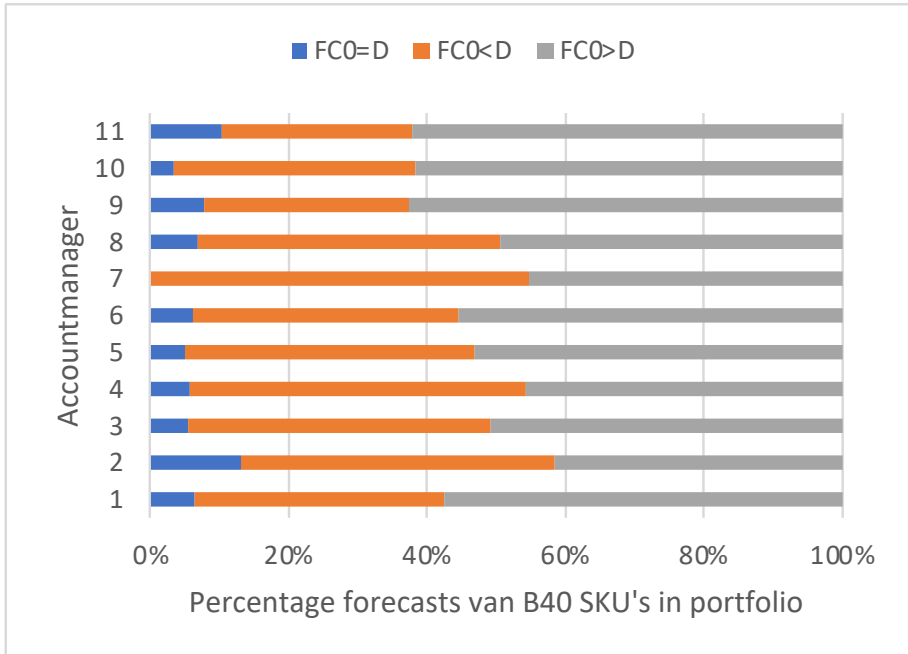
Forecastkwaliteit op individueel niveau

Om een beter beeld te krijgen over de forecastkwaliteit is met diverse accountmanagers gesproken. Gegevens over hun individuele prestaties zijn hierin meegenomen. Sommige accountmanagers gaven aan dat zij denken een lagere performance te hebben omdat hun productportfolio naar verhouding meer producten met lastig te voorspellen vraagpatronen bevat dan andere portfolio's. Daarom hebben we de verdeling in vraagpatronen per portfolio onderzocht (zie figuur 2). Het is inderdaad zo dat het merendeel van de B40 producten in de portfolio's lastiger te voorspellen vraagpatronen hebben. Slechts circa een-derde van de B40 producten heeft een smooth (relatief makkelijk te voorspellen) vraagpatroon (met uitzondering van twee uitschieters naar 40%). Echter, er zijn geen grote verschillen tussen portfolio's dus dit kan geen oorzaak zijn van een lagere performance van sommige accountmanagers.



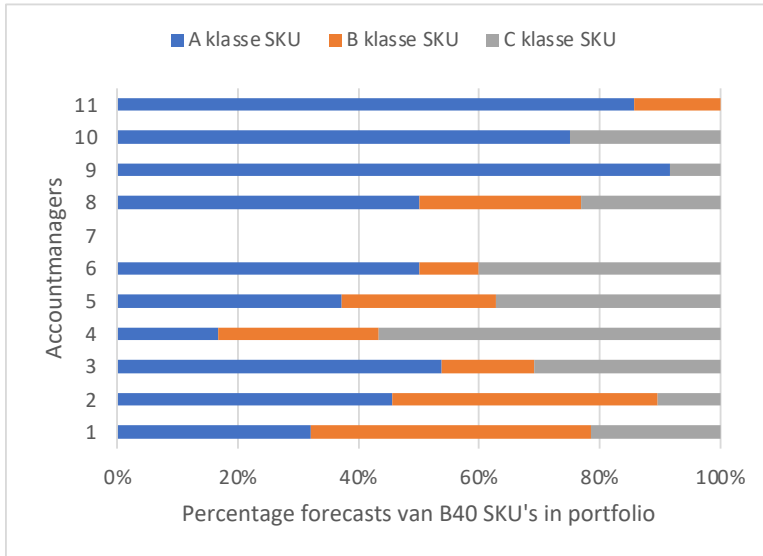
Figuur 2 Gestapelde staafdiagram met het aandeel product-vraagpatronen per portfolio

Vervolgens hebben we de accuraatheid van de voorspellingen op individueel niveau bekeken. Onderstaande figuur 3 geeft het aandeel correcte voorspellingen, overschattingen en onderschattingen weer per accountmanager. Er zijn geen grote verschillen te zien tussen accountmanagers. Opvallend is wel dat er één accountmanager is die géén voorspelling juist heeft.



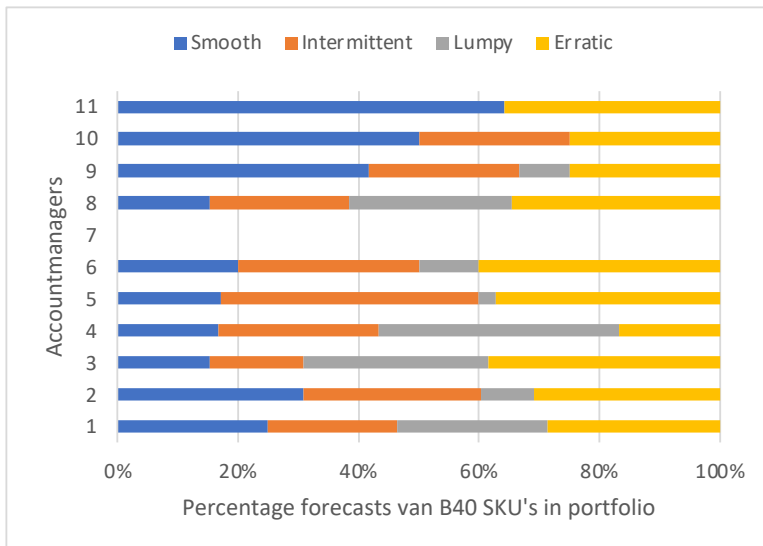
Figuur 3 Gestapelde staafdiagram met het aandeel juiste voorspellingen/overschattingen/onderschattingen per accountmanager

Om meer inzicht te krijgen in onderlinge verschillen in accurate voorspellingen, hebben we per accountmanager gekeken welk aandeel de verschillende productklassen en vraagpatronen innemen in de juiste voorspellingen, zie figuren 4 en 5 hieronder. (Het aandeel in forecasts die te laag en te hoog waren ten opzichte van de vraag is met name informatief wanneer de magnitude van de afwijking in acht wordt genomen. Hier wordt rekening mee gehouden in de analyse van de kwaliteit van de forecasts.) Er zijn onderlinge verschillen zichtbaar. De meeste accountmanagers voorspellen de minder belangrijke producten relatief vaker juist (met uitzondering van manager 9 t/m 11). Als het gaat om vraagpatronen valt te zien dat de meest accountmanagers lastiger te voorspellen vraagpatronen (dat wil zeggen, vraagpatronen met meer variatie) beter voorspellen. Het is goed mogelijk dat de accountmanagers waardevol klantcontact hebben om goede inschattingen van de vraag te kunnen maken wanneer deze lastig te voorspellen is.



Figuur 4 Gestapelde staafdiagram met het aandeel productklassen in correcte voorspellingen van accountmanagers

76



Figuur 5 Gestapelde staafdiagram met het aandeel product-vraagpatronen in correcte voorspellingen van accountmanagers

Ook op individueel niveau hebben we de forecastafwijkingen onderzocht. We hebben per productklasse en vraagpatroon de gemiddelde afwijking van de voorspelling ten opzichte van de vraag geanalyseerd. Individuele verschillen zijn met name te zien in afwijkingen van de vraag naar boven ($FC0>D$), maar dit kan mede verklaard worden door het feit dat een afwijking naar beneden niet meer dan 100% van de vraag kan afwijken (zie ook tabel 2). Geclusterde kolomdiagrammen met de resultaten per accountmanager zijn in appendix A opgenomen. Binnen dit onderzoek zijn de gevonden verschillen niet nader onderzocht. Hier ligt een mogelijkheid voor vervolgonderzoek. Wel zijn de bevindingen over individuele verschillen meegenomen bij het opstellen van advies met betrekking tot de rol van menselijke oordeel in het forecastproces.

Kwaliteit vraagvoorspellingen demand review (FC1)

Ook voor de afdeling demand review hebben we de accuraatheid van de voorspellingen in kaart gebracht. Van de 1473 door demand review uitgegeven forecasts is slechts 5,9% correct. Ook hier hebben we een onderscheid gemaakt tussen productkarakteristieken. De linkerzijde van onderstaande tabel laat het percentage juiste voorspellingen per productcategorie zien. (We willen weten van welke productcategorie de vraag relatief vaker correct wordt voorspeld, daarom corrigeren we hier voor het absolute aantal producten per categorie (en berekenen we geen percentage van het totaal aantal juiste voorspellingen).) Wat opvalt is dat A klasse producten naar verhouding het minst vaak juist voorspeld worden. Ook valt het op dat de producten die volgens hun vraagpatroon het makkelijkst te voorspellen zijn – de smooth vraagpatronen – het minst vaak juist voorspeld worden.

77

Tabel 3 Percentage juiste/overschatte/onderschatte FC1 voorspellingen per productcategorie (de dataset kent geen C producten met een smooth vraagpatroon)

	% vraagvoorspellingen correct ($FC0=D$)				% vraagvoorspellingen overschat ($FC0>D$)				% vraagvoorspellingen onderschat ($FC0<D$)			
	A	B	C	Totaal	A	B	C	Totaal	A	B	C	Totaal
Erratic	4,1%	7,5%	0,0%	4,6%	47,1%	50,0%	18,9%	43,8%	48,8%	42,5%	81,1%	51,6%
Intermittent	5,5%	5,2%	9,6%	7,5%	39,7%	34,1%	24,9%	30,2%	54,8%	60,7%	65,5%	62,3%
Lumpy	0,0%	8,3%	9,1%	8,4%	60,0%	35,5%	33,0%	35,1%	40,0%	56,2%	57,9%	56,5%
Smooth	1,4%	2,9%		1,7%	55,0%	44,3%		52,4%	43,6%	52,9%		45,8%
Totaal	2,9%	6,2%	8,4%	5,9%	50,0%	39,9%	27,4%	38,8%	47,1%	53,9%	64,2%	55,3%

Tabel 3 maakt ook een onderscheid tussen voorspellingen die de vraag overschatten en onderschatten (zie het midden en rechts van de tabel). De vraag naar producten wordt vaker onderschat dan overschat (38,8% versus 55,3%, respectievelijk). Dit geldt voor de verschillende productklassen (met uitzondering van A klasse producten) en vraagpatronen. Verder is te zien dat als de vraag naar producten wordt overschat, dit relatief het vaakst gebeurt voor de meest belangrijke producten (A klasse). Als de vraag wordt onderschat zijn het relatief vaker B en C klasse producten. Als we naar het vraagpatroon kijken, zien we dat smooth vraagpatronen, hoewel die het meest stabiel zijn, relatief het vaakst worden overschat. Deze bevinding is in lijn met eerdere bevindingen dat juist stabiele (smooth) vraagpatronen relatief vaker worden overschat. Interessant genoeg worden lumpy vraagpatronen, die de meeste variatie kennen, niet het vaakst overschat of onderschat.

Van de foute voorspellingen hebben we ook weer de forecastafwijking geanalyseerd. Onderstaande tabel 4 geeft de resultaten weer. Opvallend is dat de gemiddelde positieve afwijking van de vraag ($FC0 > D$) het grootst is voor A en B klasse producten terwijl de gemiddelde negatieve afwijking ($FC0 < D$) het grootst is voor C klasse producten. Met andere woorden, belangrijkere producten worden, relatief tot de vraag gezien, het meest overschat en minder belangrijke producten het meest onderschat. Verder is te zien dat de gemiddelde positieve en negatieve afwijking het grootst is voor producten met een lumpy vraagpatroon. Dit vraagpatroon is dan ook wel relatief het lastigst te voorspellen (gegeven de variëteit in timing en hoeveelheid). De gemiddelde afwijking is het kleinst voor producten met een smooth (stabiel) vraagpatroon.

78

Tabel 4 Gemiddelde kwaliteit van $FC0$ per productcategorie

	Gemiddelde procentuele afwijking van overschatte vraagvoorspellingen ($FC0 > D$)				Gemiddelde procentuele afwijking van onderschatte vraagvoorspellingen ($FC0 < D$)			
	A	B	C	Totaal	A	B	C	Totaal
Erratic	272%	4985%	1243%	2209%	-41%	-51%	-68%	-51%
Intermittent	149%	182%	345%	245%	-50%	-57%	-75%	-66%
Lumpy	191674%	19173%	4578%	24334%	-80%	-80%	-83%	-82%
Smooth	99%	69%		93%	-36%	-44%		-38%
Totaal	7378%	5888%	2375%	2296%	-42%	-60%	-77%	2296%

Als we tabel 1 en 2 met elkaar vergelijken, valt ook nog iets op. Het vaakst of het minst vaak overschatten of onderschatten van de vraag, zegt niet altijd iets over de procentuele afwijking van de vraag. Dat wil zeggen, niet als we kijken naar het vraagpatroon. Hoewel de vraag naar producten met een smooth vraagpatroon het vaakst wordt overschat, is de gemiddelde afwijking wel het kleinst. Hoewel lumpy vraagpatronen het minst vaak worden overschat, is de gemiddelde procentuele afwijking per overschatting wel het grootst.

Kwaliteit wijzigingen vraagvoorspelling door demand review

We hebben ook de wijzigingen die demand review in FC1 maakt van de op productniveau geaggregeerde FC0's bekeken. We bekijken welke voorspellingen worden aangepast, wat voor soort aanpassingen er worden gedaan en wat de kwaliteit van de aanpassingen is.

In totaal is 3,5% van de FC0 voorspellingen door demand review gewijzigd. Het percentage wijzigingen per productcategorie is weergegeven in tabel 5 hieronder. De meeste wijzigingen zijn gemaakt voor A klasse producten. Dit ligt voor de hand omdat het demand review team daar extra aandacht aan besteedt vanwege het belang van deze producten voor de organisatie. Deze bevinding is ook in lijn met eerder onderzoek (Khosrowabadi et al., 2022). Hoewel deze voorspellingen wel gewijzigd worden, hebben we hierboven gezien dat in absolute zin deze categorie wel het minst vaak juist wordt voorspeld. De kwaliteit van de gemaakte aanpassingen wordt hieronder verder onderzocht. Ook is in tabel 5 te zien dat producten met een smooth vraagpatroon en met een erratic vraagpatroon het vaakst worden gewijzigd.

79

Tabel 5 Percentage gewijzigde vraagvoorspellingen in FC1 per productcategorie

	A	B	C	Totaal
Erratic	7,6%	4,2%	0,0%	5,2%
Intermittent	4,1%	1,2%	0,4%	1,2%
Lumpy	0,0%	2,5%	1,5%	1,8%
Smooth	9,6%	0,0%		7,3%
Totaal	7,7%	2,1%	0,8%	3,5%

We hebben ook de richting van de aanpassingen onderzocht. De meeste aanpassingen betrof het naar beneden bijstellen van de geaggregeerde FC0 voorspelling (84,3% van de wijzigingen). In verhouding tot het totaal aantal voorspellingen per categorie, werden A klasse producten met een smooth of erratic vraagpatroon relatief het vaakst naar boven aangepast. Het zijn ook dit type producten die relatief het vaakst naar beneden worden aangepast (en in absolute zin ook vaker naar beneden dan boven).

Vervolgens hebben we geanalyseerd of de gemaakte wijzigingen juist waren. Een kleine meerderheid van de aanpassingen zijn juist, 59%. De onderstaande tabel 6 toont het aandeel correcte aanpassingen ten opzichte van het totaal aantal aanpassingen per productcategorie. Te zien is dat aanpassingen van voorspellingen van C klasse producten en van producten met een erratic, lumpy en smooth vraagpatroon het vaakst juist zijn. Aanpassingen van vraagvoorspellingen van producten met een intermittent vraagpatroon zijn verhoudingsgewijs het vaakst fout.

Tabel 6 Percentage juiste aanpassingen per productcategorie (ten opzichte van het aantal aanpassingen per categorie)

	A	B	C	Totaal
Erratic	61,5%	100,0%		72,2%
Intermittent	33,3%	0,0%	0,0%	16,7%
Lumpy		33,3%	100,0%	66,7%
Smooth	57,1%			57,1%
Totaal	56,8%	60,0%	75,0%	58,8%

80

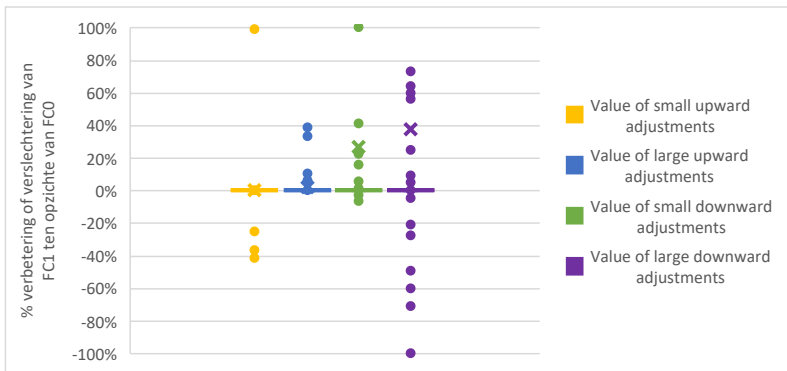
Ook hebben we onderzocht hoe goed de gemaakte aanpassingen zijn, dat wil zeggen, in hoeverre ze de voorspellingen hebben verbeterd. Om de kwaliteit van de aanpassingen te bepalen berekenen we hoeveel dichter FC1 bij de vraag ligt (of hoeveel er verder van af) dan FC0: $(FC0_{E/D} - FC1_{E/D})$. Een positieve waarde duidt op een correcte aanpassing (de voorspelfout van FC0 is groter dan de vraag) en een negatieve waarde duidt op een incorrecte aanpassing. Hoe groter de waarde, hoe groter de verbetering/verslechtering van de aanpassing in verhouding tot de gerealiseerde vraag.

Vervolgens maken we een onderscheid tussen grote en kleine aanpassingen naar boven en beneden. Dit doen we door de aanpassingen te ordenen op magnitude en vervolgens aanpassingen naar boven (beneden) in twee gelijke 50-50 te verdelen. Grote en kleine verbeteringen/verslechtingen zijn op eenzelfde wijze onderscheiden (aanpassingen ordenen en verbeteringen (verslechtingen) 50-50 verdelen). Daarna kunnen we bepalen hoe goed of slecht kleine en grote aanpassingen naar boven en beneden zijn. Het percentage aanpassingen per categorie is weergegeven in onderstaande tabel 7. De resultaten zijn ook gevisualiseerd in onderstaande boxplot in figuur 6.

Tabel 7 Percentage kleine/grote verbetering/verslechtering per richting en omvang van de aanpassing

		Relatieve waarde van de aanpassing				
		Fout (-23% tot 0)	Goed (0 tot 29%)	Erg goed (> 29%)	Totaal	
Aanpassing naar boven	Klein (0-5250)	5,9%			2,0%	7,8%
	Groot (>5250)			3,9%	3,9%	7,8%
Aanpassing naar beneden	Klein (-1828 tot 0)		17,6%	17,6%	7,8%	43,1%
	Groot (< -1828)	13,7%	3,9%	7,8%	15,7%	41,2%
	Totaal	19,6%	21,6%	29,4%	29,4%	100%

Te zien is dat aanpassingen naar boven vaker goed zijn dan niet goed. Met name kleine aanpassingen naar boven pakken erg negatief uit, grote aanpassingen pakken juist goed uit. Aanpassingen naar beneden pakken vaker goed uit dan slecht. Het is interessant om te zien dat kleine aanpassingen naar beneden merendeels een kleine verbetering of verslechtering opleveren terwijl grote aanpassingen naar beneden relatief vaker een grote verbetering of verslechtering opleveren.



Figuur 6 Boxplot met de spreiding van de mate van verbetering/verslechtering van FC1 ten opzichte van FCO (in verhouding tot de vraag) (noot: omwille van leesbaarheid is de boxplot is ingezoomd op aanpassingen de tot 100% verbetering/verslechtering opleverden)

Discussie en advies

Uit de analyse van het forecastproces bij TDF blijkt dat accountmanagers elke maand veel vraagvoorspellingen moeten maken, sommigen wel meer dan 300. Dit is een tijdrovend proces waarbij keuzes gemaakt moeten worden qua focus op klantvraag-voorspellingen. Bovendien gebeurt het voorspellen handmatig, zonder gebruik van een geautomatiseerd systeem die met behulp van statistische analyses voorstellen kan doen. Dit versterkt de druk op de accountmanagers. Zij maken voorspellingen aan de hand van eigen inschattingen, ervaring, marktkennis en klantcontact.

De kwaliteit van de vraagvoorspellingen van accountmanagement en demand review bevat ruimte voor verbetering. Zo'n 7,0 en 5,9% van de voorspellingen, respectievelijk, zijn accuraat. (De forecastafwijking kunnen we niet één-op-één vergelijken omdat de voorspellingen van accountmanagement op klantvraag-niveau zijn en die van demand review zijn geaggregeerd op SKU niveau.) We zien een aantal tendensen. Zowel accountmanagers als demand review schatten de vraag naar A klasse producten, die het belangrijkste zijn voor TDF, over het algemeen niet het beste in. Dit is opvallend omdat de afdelingen daar over het algemeen wel de meeste nadruk op leggen vanwege het belang van deze producten. Verder zien we het overschatten van C klasse producten bij accountmanagement niet terug bij demand review. Na aanpassingen wordt bij demand review met name de vraag naar A klasse producten overschat.

82

Ook valt op dat beide afdelingen juist producten met een wat lastiger vraagpatroon goed voorspellen. Een mogelijke verklaring ligt in het feit dat accountmanagers nauw contact hebben met hun klanten. Waar het maken van een inschatting lastig is, is klantinformatie belangrijk om een juiste forecast te maken. Mogelijk hebben accountmanagers minder nauw contact bij makkelijker te voorspellen producten (met een smooth vraagpatroon) omdat ze dan meer vertrouwen op hun eigen kennis. Verder zien we de in de theorie genoemde tendens dat stabielere vraagpatronen relatief vaker overschat worden (Eroglu & Croxton, 2010; Kremer et al., 2011), terug bij demand review.

Demand review past slechts een klein deel van de voorspellingen van accountmanagement aan en richt zich met name op de belangrijkste producten. Dit is in lijn met resultaten van eerder onderzoek (Khosrowabadi et al., 2022). Verreweg de meeste aanpassingen werden gedaan naar beneden. Dit kan deels verklaard worden door beleid vanuit het moederbedrijf van TDF. Demand review wordt gestimuleerd om niet al te positieve voorspellingen uit te geven doordat het moederbedrijf stuurt op het matchen van productie met de vraag met beperking van overproductie. Door voorspellingen naar beneden bij te stellen, worden productie- en dus voorraadniveaus gedrukt. Daarmee wordt het waarschijnlijker dat er geen overproductie plaatsvindt. Het naar beneden bijstellen van de voorspellingen is een vorm van risico-aversie (Díaz & Esparcia, 2019).

Om het vraagvoorspelproces bij TDF te verbeteren, stellen we een aantal adviezen voor. We bespreken ze hier achtereenvolgend.

Advies 1: Automatiseren waar het kan, interveniëren waar nodig

Om de accountmanagers werk uit handen te geven en te zorgen dat zij tijd hebben om goed naar belangrijke voorspellingen te kijken, adviseren we om het voorspelproces deels te automatiseren. Deels, omdat voor sommige maar niet alle producten een geautomatiseerd systeem beter presteert (Syntetos et al., 2005). Omdat de geschikte rekenmethode afhangt van het vraagpatroon is het advies daarbij om de gebruikte rekenmethode ook te diversifiëren.

We adviseren om vraagvoorspellingen van producten met smooth vraagpatroon volledig te automatiseren. Vanwege de relatief lage variatie in timing en hoeveelheid, is de vraag van deze producten redelijk goed met statistische voorspelmethodes te voorspellen. Dit is waardevol omdat demand review dit type vraagpatroon het slechtste voorspelt (zie tabel 3). (Daarbij moet gezegd worden dat de gemiddelde afwijking van de vraag wel het kleinst is, deze voorspelfouten hebben relatief gezien wel de kleinste impact (zie tabel 4).) Demand review maakt ook de meeste aanpassingen aan voorspellingen met een smooth demand pattern (zie tabel 5) en in slechts een kleine meerderheid van de gevallen zijn die aanpassingen juist (zie tabel 6). Het automatiseren van voorspellingen van smooth producten neemt demand review dus ook controle taken uit handen.

83

Om te kunnen adviseren welke rekenmethode het beste gebruikt kan worden om de vraag naar smooth producten te schatten, zijn de in de theorie voorgestelde rekenmethodes Moving Average en Exponential Smoothing zijn getest. Volgens beide methodes zijn vraagvoorspellingen gemaakt en die zijn vergeleken met de voorspellingen van de accountmanagers en de gerealiseerde vraag. Omdat een Moving Average in meer dan 50% van de gevallen een betere inschatting maakt dan Exponential Smoothing en het menselijk oordeel, wordt voorgesteld om de voorspellingen van klantvraag van smooth producten volledig te automatiseren volgens een Moving Average met 6 periodes (Mattson, 2010).

We adviseren om voor intermitterend vraagpatronen statistische voorspelmethodes in te zetten die wel door accountmanagers en demand review aangepast kunnen worden. Voor deze vraagpatronen hebben we namelijk de voorspellingen van de accountmanagers vergeleken met die van de Crostons methode en Exponential Smoothing 'met' en 'zonder' menselijk oordeel. (Ongeveer de helft van de producten met dit vraagpatroon is in deze analyse meegenomen omdat de andere helft slechts afzet had in één of twee maanden en dan werken de voorspelmethodes niet.) Met menselijk oordeel wil zeggen dat in periodes waar de afzet en de voorspelling 0 waren, de statistische voorspelling ook op 0 is gezet. Accountmanagers weten, in tegenstelling tot een systeem, vanuit hun marktkennis en

klantcontact namelijk in de meeste gevallen wanneer geen afzet komt. De maanden waar wel afzet is, wordt de voorspelling van de rekenmethode gebruikt.

Beide voorspelmethodes werken beter in combinatie met menselijk oordeel. Maar omdat bij Exponential Smoothing vraag van de afgelopen drie maanden wordt beschouwd, is de kans op 'nee-verkopen' groter dan bij de Crostons methode. Daarom wordt de Crostons methode in combinatie met menselijk oordeel geadviseerd. De accountmanager krijgt de Crostons voorspelling te zien en kan deze op basis van markt- en klantkennis nog aanpassen. Producten met een intermittent vraagpatroon worden door demand review weliswaar niet het slechtste ingeschat en de gemiddelde afwijking van de vraag is relatief minder groot dan bij smooth producten, maar de gemaakte aanpassingen aan deze voorspellingen zijn relatief gezien wel het slechtst. Een ondersteunend systeem dat met een voorstel komt, kan hier hulp bieden.

Producten met een erratic vraagpatroon kunnen met een statistische methode voorspeld en maar dit vraagt expertise en veel rekenkracht. Omdat slechts 10% van het productportfolio bij TDF dit vraagpatroon kent, wordt hier geadviseerd menselijk oordeel in te zetten. Producten met een lumpy vraagpatroon zijn dermate slecht te voorspellen dat ook hier wordt geadviseerd menselijk oordeel in te zetten (Ragnerstam, 2015; Yaniv & Hogarth, 1993). Het uitgaan van menselijk oordeel bij lumpy en erratic vraagpatronen sluit in die zin aan bij de geobserveerde kwaliteit van voorspellingen: deze patronen worden door demand review het vaakst juist geforecast. De gemiddelde afwijking van de vraag is wel het hoogst, maar dit is ook inherent aan de relatief grote mate van variabiliteit in gevraagde hoeveelheid bij dit type vraagpatronen. Daarnaast zijn de gemaakte aanpassingen door demand review bij deze typen vraagpatronen het vaakst juist (in 2/3 of meer van de gevallen, zie tabel 6).

Advies 2: een 'bayesian' benadering ter ondersteuning menselijk oordeel

Het tweede advies is om een 'bayesian benadering' te implementeren voor de voorspellingen waar rekenmethodes ingezet gaan worden. Dit houdt in dat voorspellers informatie te zien krijgen over de waarschijnlijkheid van de systeemvoorspelling (hoe groot de kans is dat de werkelijke vraag hoger of lager gaat zijn dan de door de voorspelmethode voorspelde vraag). Daarnaast wordt informatie gerapporteerd over of de eigen aanpassingen aan de systeemvoorspelling van de afgelopen periode juist zijn. Dit leidt tot meer inzichten in de eigen voorspelling en de voorspelmethode. De voorspeller leert zodoende de methode kennen, wat leidt tot een hogere betrouwbaarheid van de voorspellingen (Zellner et al., 2021).

Advies 3: min-max methode voor voorraadbeheer

Het derde advies is om de min-max methode te implementeren voor het beheren van de voorraad van A klasse producten. (De voorspelmethode kan voor A klasse producten op de achtergrond gebruikt blijven worden ter ondersteuning om de 'max' vast te stellen voor een productiecycclus van een product met een smooth vraagpatroon.) Bij TDF zijn dit producten met smooth, intermittent en erratic vraagpatronen. Aan de hand daarvan kunnen ook automatisch productieniveaus vastgesteld worden. Deze oplossing neemt precies een deel van de voorspelregels weg waar de forecastkwaliteit relatief laag is. Demand review schat de vraag naar A klasse producten namelijk het minst vaak juist in. Ook worden voorspellingen van A klasse producten het vaakst gewijzigd maar zijn de wijzigingen relatief het minste vaak juist.

Het wordt geadviseerd om voorraadniveaus voor producten afzonderlijk te bepalen omdat ieder product een ander vraagpatroon en productietijd heeft (Fokkens, 2011). Een geschikte min-max methode is een stochastisch model in combinatie met de bestelpuntmethode van Visser en Van Goor (2015). In overleg met TDF wordt geadviseerd productieniveaus te bepalen aan de hand van de huidige voorraadstand met een servicegraad van 95%. (Bij een hogere servicegraad neemt de voorraad namelijk exponentieel toe.) Het periodiek controleren van de min-max waarden is belangrijk om recente vraagschommelingen op te vangen.

85

Advies omtrent aanpassingen demand review

Uit de analyse van de kwaliteit van de aanpassingen van demand review (zie tabel 7) komt naar voren dat aanpassingen naar boven vaker goed zijn dan niet goed. Maar met name grotere aanpassingen pakken goed uit. Aanpassingen naar beneden zijn net vaker goed dan niet goed (circa 60-40). Kleine en grote aanpassingen kunnen een kleine of grote verbetering opleveren. Het is daarom verstandig om grotere aanpassingen met zorg te overwegen.

Conclusie

Het correct voorspellen van klantvraag is belangrijk om betrouwbare plannings te maken voor - en de juiste beslissingen te nemen over - voorraden, productie en inkoop. Hiermee kan beter worden ingespeeld op de klantvraag en kunnen kosten gereduceerd worden. Het maken van goede vraagvoorspellingen is echter niet eenvoudig. Geautomatiseerde systemen kunnen voorspellers ondersteunen door statistische voorspellingen te maken. In de praktijk worden die systemen echter niet altijd gebruikt of daarbovenop vindt nog een menselijke beoordeling plaats. Menselijke aanpassingen kunnen voorspellingen verbeteren omdat mensen over impliciete kennis beschikken, maar aanpassingen zijn niet altijd waardevol.

Bij de inrichting van een forecastproces moet daarom nagedacht worden over de rol van menselijke beoordelingen en interventies. In dit project hebben we de kwaliteit van het forecastproces bij TDF onderzocht en gekeken waar en hoe het forecastproces geautomatiseerd kan worden en welke rol menselijk oordeel daarin zou moeten hebben. Er wordt geadviseerd om het forecastproces te automatiseren waar mogelijk, en het menselijk oordeel in te zetten waar nodig.

Bij TDF maken accountmanagers maandelijks veel voorspellingen en deze worden handmatig gedaan. Hierin is ruimte voor verbetering. We hebben gezien dat de belangrijkste producten (A klasse) niet het beste voorspeld worden en dat juist de vraag naar producten met een niet-stabiel vraagpatroon beter voorspeld worden. Verder wordt slechts een klein deel van de voorspellingen aangepast, waarvan maar een kleine meerderheid een juiste aanpassing is gebleken.

Om voldoende tijd te hebben voor belangrijke of lastige voorspellingen wordt geadviseerd om voorspellingen van producten met een smooth vraagpatroon geheel te automatiseren en die van producten met een intermitterend vraagpatroon te automatiseren maar ruimte te laten voor systeemaanpassingen. Vanwege de complexiteit van erratic en lumpy vraagpatronen wordt geadviseerd die handmatig te voorspellen. De accuraatheid van de voorspellingen van deze vraagpatronen door demand review zijn ook het hoogst. Verder wordt bij automatisering een 'bayesian' benadering geadviseerd waarbij voorspellers informatie zien over de over- en onderschattingskans van de methode en de kwaliteit van eerdere aanpassingen. Het toepassen van een min-max methode voor voorraadbeheer van de A klasse producten zou helpen het aantal voorspelregels verder te reduceren en de werkdruk van voorspellers te verlichten. Demand review wordt verder geadviseerd om met name grote aanpassingen naar beneden om grotere aanpassingen met zorg te overwegen.

86

References

- Boone, T., Ganeshan, R., Jain, A., & Sanders, N. R. (2019). Forecasting sales in the supply chain: Consumer analytics in the big data era. *International Journal of Forecasting*, 35, 170–180.
- De Baets, S., & Harvey, N. (2018). Forecasting from time series subject to sporadic perturbations: Effectiveness of different types of forecasting support. *International Journal of Forecasting*, 34, 163–180.
- Díaz, A., & Esparcia, C. (2019). Assessing risk aversion from the investor's point of view. *Frontiers in Psychology*, 10(JULY). <https://doi.org/10.3389/fpsyg.2019.01490>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>

- Erasmus University Rotterdam. (2018). Data science & machine learning: het gebruik van state of the art voorspelmodellen in data science. Geraadpleegd op 3 juni 2022, van "https://www.eur.nl/sites/corporate/files/Data_Science_Januari_2018.pdf
- Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*, 116–133.
- Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*, 3–23.
- Fokkens, D. (2011). *Kwantitatief voorraadbeheer magazijnartikelen MTS*. Universiteit Twente.
- Freppele.com, (z.d.). *Demand classification: Why forecastability matters*. Frepple.com. Geraadpleegd op 3 juni 2022, van https://frepple.com/blog/demand-classification/
- Fu, W., Chien, C., & Lin, Z. (2018). A hybrid forecasting framework with neural network and time-series method for intermittent demand in semiconductor supply chain. *IFIP International Conference on Advances in Production Management Systems (APMS)*, 6–72.
- Jiang, P., Liu, X., Huang, Y., & Yuan, Y. (2017). An adaptive data-driven model for erratic demand forecasting. *IFAC OapersOnLine*, 9020–9025.
- Khosrowabadi, N., Hoberg, K., & Imdahl, C. (2022). Evaluating human behaviour in response to AI recommendations for judgmental forecasting. *European Journal of Operational Research*, *303*, 1151–1167.
- Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, *57*(10), 1827–1843. https://doi.org/10.1287/mnsc.1110.1382
- Mattson, S. (2010). *Effektivare materialstyrning – En handbok för att lyckas*. Permatron.
- Ragnerstam, E. (2015). *How to calculate forecast accuracy for stocked items with a lumpy demand - A case study at Alfa Laval*. Malardalens Hogskola Eskilstuna Vasteras.
- Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, *31*, 511–522.
- Siemsen, E., & Aloysius, J. (2019). *Supply chains analytics and the evolving work of supply chain managers*. https://doi.org/10.13140/RG.2.2.15396.30081
- Syntetos, A. A., Boylan, J. E., & Croston, J. D. (2005). On the categorization of demand patterns. *Journal of the Operational Research Society*, *56*(5), 495–503. https://doi.org/10.1057/palgrave.jors.2601841

Talele, J. (2020, 14 januari). *Croston model: Forecasting Intermittent demand data (Time Series analysis)*. Medium. Geraadpleegd op 2 juni 2022, van <https://juileetalele.medium.com/croston-model-forecasting-intermittent-demand-data-time-series-analysis-6f3a2bb154b>

Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, 29, 34–243.

Visser, H. M., & Van Goor, A. R. (2015). *Werken met logistiek* (7e editie). Noordhoff Uitgevers.

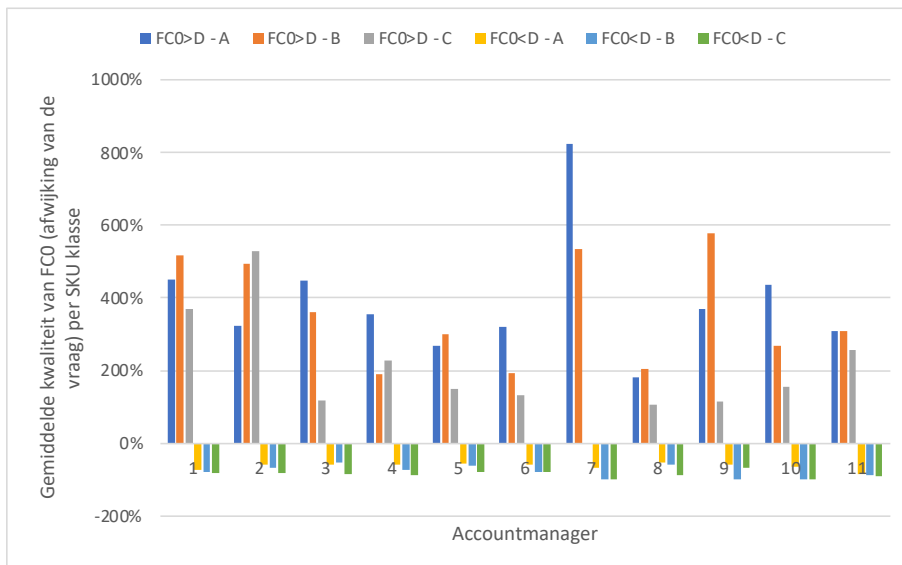
Yaniv, I., & Hogarth, R. M. (1993). Judgmental versus statistical prediction: Information asymmetry and combination rules. *Psychological Science*, 4(1), 58–62. <https://doi.org/10.1111/j.1467-9280.1993.tb00558.x>

Zellner, M., Abbas, A. E., Budescu, D. V., & Galstyan, A. (2021). A survey of human judgement and quantitative forecasting methods. *Royal Society Open Science*, 8(2). <https://doi.org/10.1098/rsos.201187>

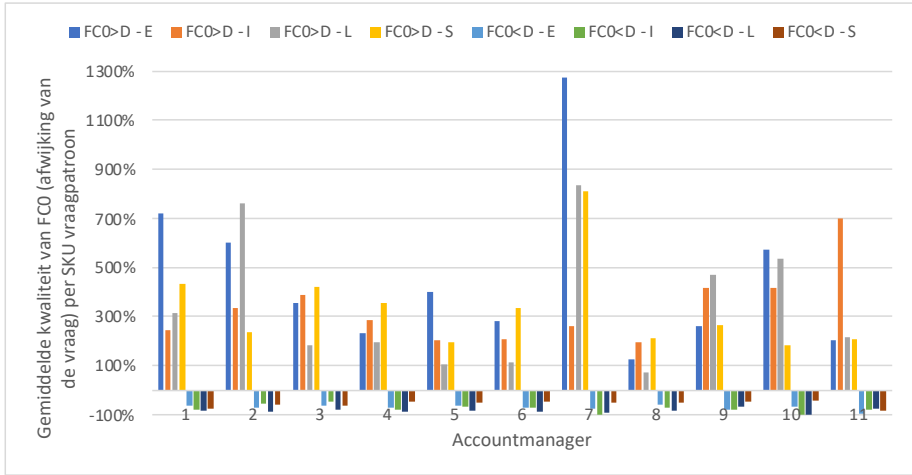
Appendix

Onderstaande geclusterde kolomdiagrammen geven per accountmanager per SKU klasse (figuur A) en vraagpatroon (figuur B) de gemiddelde forecastafwijking van de vraag. Te zien is dat met name in de voorspellingen die de vraag overschatten grotere verschillen in de gemiddelde overschatting per productklasse en vraagpatroon te zien zijn.

88



Figuur A Geclusterde kolomdiagram met gemiddelde kwaliteit FC0 per accountmanager per productklasse



Figuur B Geclusterde kolomdiagram met gemiddelde kwaliteit FCO per accountmanager per vraagpatroon